



Python para Data Science

Prof. Jorge Zavaleta

Prof. Sergio Serra

Prof. Macário Costa



PET-SI

Sobre o Curso de Extensão



Apresentar os fundamentos de Python para Data Science



Preparar e capacitar alunos e professores para manipular Big Data



Aplicar as diferentes técnicas de aprendizado de máquina para manipular dados de diversos formatos



Usar o aprendizado de máquina para resolver problemas práticos



Visualizar o comportamento dos dados e fazer uma análise dos resultados



Sobre os Professores



PET-SI



Prof. Sergio Manuel Serra da Cruz

Prof. Raimundo Macário Costa

Prof. Jorge Zavaleta (Externo à UFRJ)





PET-SI

Ementa do Curso





PET-SI

Módulo 1

1. Ambiente de Programação: python 3, jupyter notebook, pacotes e github
2. Python I: tipos de dados, sequências e operações, estruturas de controle e repetição
3. Prática dos conteúdos estudados: construindo e operando listas e strings.





PET-SI

Módulo 2

1. Numpy: array, slicing, fancy index, copy and view
2. Pandas: dataframes, series, index, Pandas I/O (csv, json, excell)
3. Prática dos conteúdos estudados: Processando e extraíndo informações de arquivos csv





PET-SI

Módulo 3

1. Numpy e Matplotlib: operações com array, broadcasting, construção de gráficos usuais
2. Pandas II: limpeza de dados e estatísticas básicas
3. Prática dos conteúdos estudados: manipulando dados do ensino médio





PET-SI

Módulo 4

1. Scikit learn: introdução a mecanismos de clustering, regressão e classificação
2. Prática dos conteúdos estudados: clusterizando e predizendo o desempenho de escolas do ensino médio.





PET-SI

Avaliação

- Grupos de X: Usar Deepnote, Jupyter, JupyterLab, Colab para fazer os exercícios e projetos.
- Lista de Exercícios
- Projeto Final de Análise de dados com *datasets* diferentes.
- Datas de entrega: (xx Julho) por confirmar!!!



O que é Data Science



PET-SI

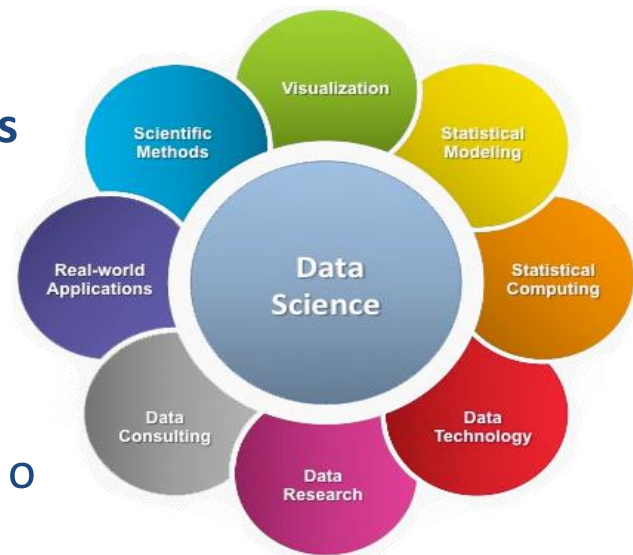




PET-SI

O que é Data Science?

- Ciência de dados é o estudo dos dados.
- Ciência que **analisa dados**, combinando a **estatística** com **aprendizado de máquina**.
- Tenta responder ao desafio que o **Big Data** apresenta.



Fonte da Imagem:

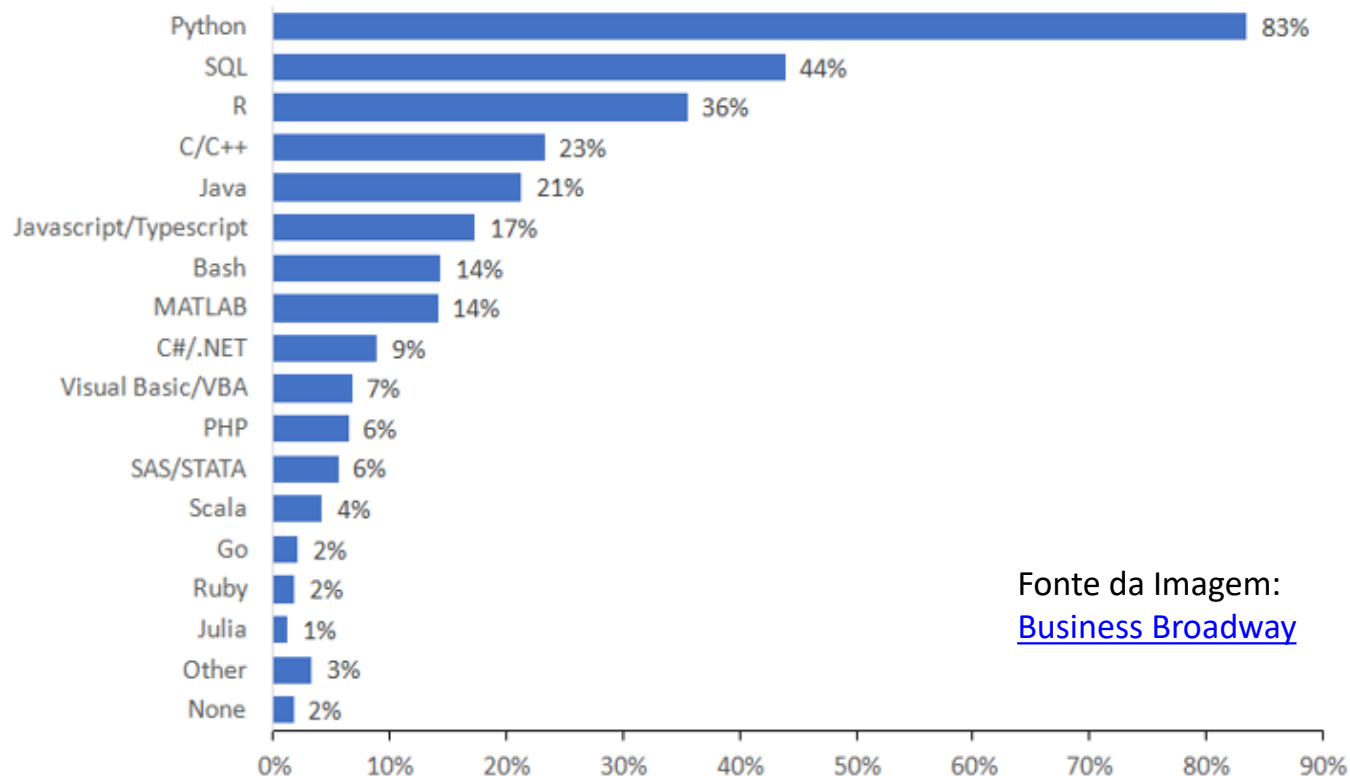
<http://www.kiwidatascience.com/>





PET-SI

Linguagens de Programação



Fonte da Imagem:

[Business Broadway](#)





PET-SI

Ferramentas para Data Science

R

usado para
análise
estatística

BigML

ML para DS de fácil
uso

Tableau

simplicifica dados
brutos em formato
compreensível

SAS

facilita o análise,
reportes, e
mineração de
dados

MatLab

usado para
processamento de
dados

Ggplot2

usado para
Visualização gráfica

MatplotLib

geração de gráficos
com análise de
dados

Jupyter

ajuda a
transformar e
visualização de
dados





PET-SI

Configurando o ambiente de trabalho

Configuração do Ambiente



O curso será feito em **JUPYTER**
Pode ser usado qualquer ambiente de programação
As aulas estão disponíveis no [GitHub](https://github.com/zavaleta/Data_Science_UFRSJ)

https://github.com/zavaleta/Data_Science_UFRSJ

Será usado o Python 3 através do ambiente ANACONDA

Configuração do Ambiente

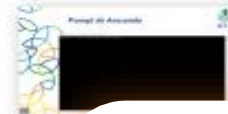


Instalação para Usuários **EXPERIENTES**:

O Python poderá ser instalado usando "Python package manager", usando o comando, **pip**, ao invés do ambiente Anaconda.

Abrir o prompt de comandos e usar:

pip3 install jupyter



Configuração do Ambiente



Instalação para Usuários **NOVOS**:

Para usuários **novos** recomenda-se instalar o ambiente **Anaconda**.

✓ <https://www.anaconda.com/distribution/>

Acessar o site do **JUPYTER** e seguir o roteiro para a instalação

<https://jupyter.org/>





PET-SI

Configuração do Ambiente

O curso será feito em **JUPYTER**

Pode ser usado qualquer ambiente de programação

As aulas estão disponíveis no [GitHub](#)

[https://github.com/zavaleta/Data Science UFRRJ](https://github.com/zavaleta/Data_Science_UFRRJ)

Será usado o Python 3 através do ambiente ANACONDA





PET-SI

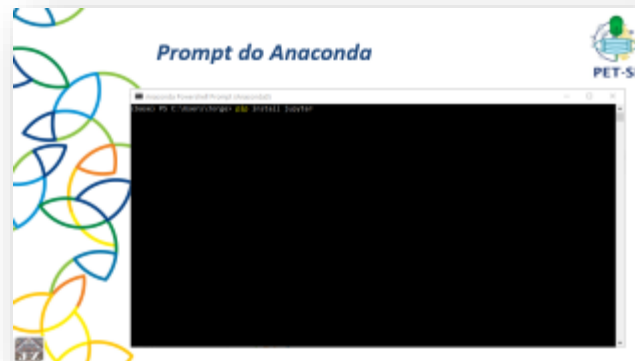
Configuração do Ambiente

Instalação para Usuários **EXPERIENTES**:

O Python poderá ser instalado usando “*Python package manager*”, usando o comando, **pip**, ao invés do ambiente Anaconda.

Abrir o prompt de comandos e usar:

pip3 install jupyter



Prompt do Anaconda



PET-SI

```
Anaconda Powershell Prompt (Anaconda3)
(base) PS C:\Users\Jorge> pip install jupyter
```





PET-SI

Configuração do Ambiente

Instalação para Usuários **NOVOS**:

Para usuários **novos** recomenda-se instalar o ambiente Anaconda.

✓ <https://www.anaconda.com/distribution/>

Acessar o site do **JUPYTER** e seguir o roteiro para a instalação

<https://jupyter.org/>





PET-SI



Jupyter Notebook



JupyterLab



Colab



Deepnote





Logout

Aprendizado de Máquina na Prática

Prof. Dr. Jorge Zavaleta

Requisitos de Sistema para usar o "NOTEBOOK"

- Python (3.7 ou 2.7)
 - Anaconda
- Jupyter Notebook
 - Desktop
 - Nuvem
- JupyterLab
 - Desktop
- Colab (Nuvem)
 - <https://colab.research.google.com/>

titulo

<https://jupyter.org/>

Bibliotecas essenciais

Project Jupyter | Home x

Project Jupyter | Home x

Google Drive/Colab N x

Aprendizado_maquin x

JupyterLab x

+ x

localhost:8888/lab

☆

☆

🔒

Não sincronizando

👤

⋮

File

Edit

View

Run

Kernel

Tabs

Settings

Help

+ 📁 ⬆ ↺

/ Google Drive / Colab Notebooks /

Name

📁 NetCDF4_python2...

📄 p2_indices.ipynb

📄 Palestra_niteroi_0...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Data_Scie...

📄 Python_Oficina_01...

📄 Redes_Neura...

📄 Spi_mlp.ipynb

📄 spi_teste.ipynb

📄 style_file0.html

📄 TensorFlow_2_0_+...

📄 Tensorflow_classifi...

📄 Teste_formularios....

📄 Teste_importar.ipy...

📄 teste_oficina.ipynb

🔗 Launcher x

📄 Classe_data_cience x

📄 MSS_Lab_Aula_0C ●

📄 Python_Data_Scie ●

📄 Redes_Neura...

📄 FGV_Workshop_F x

📄 + ✂ 📄 📄 ▶ ■ ↺ ▶▶

Markdown ▾

Python 3 ○

Redes Neura...

Perceptron

O **perceptron** é um **classificador binário linear simples**. É uma aplicação que tem como entrada um vetor x para uma saída binária $f(x)$. Dado um vetor de pesos w , a função de classificação do perceptron é:

$$f(x) = 1 \text{ se } w \cdot x + b > 0$$

or

$$f(x) = 0, \text{ outra - opção}$$

Aqui o valor do b bias é o responsável por dividir o plano.

Exemplo prático

[]:

```
# bibliotecas necessarias

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
from sklearn.preprocessing import LabelEncoder, StandardScaler
# importante o perceptron
from sklearn.linear_model import Perceptron
from sklearn.neural_network import MLPClassifier
```

0

\$

6

Python 3 | Idle

Saving completed

Mode: Command

Ln 1, Col 1

Redes_Neura...

aula.ipynb

×

Aplicando zoo de aumento

+ Code + Text

```
[27] 1 # plot train pictures
      2 plotImages(sample_training_images[:8])
```



```
[28] 1 # activation: relu, sigmoid, softmax
```

```
2 model = Sequential([
3     Conv2D(16, 3, padding='same', activation='relu',
```


Gmail

Mix Cumbia Cover DJ C3SAR

Google Tradutor

UFRRJ Python para Data Science

beta.deepnote.com/project/1cae0014-04c0-48f1-90e2-f72240454df1#%2Fintroducao.ipynb

Share

Project

Environment

Comments

Add files

Connection

Terminal

data

images

introducao.ipynb

modulo_dois.ipynb

modulo_quatro.ipynb

modulo_tres.ipynb

modulo_um.ipynb

0 global variables

Ready

Run notebook

Restart notebook

UFRRJ

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO

PET-SI

Python para Data Science

Prof. Jorge Zavaleta

ESCOLA DE EXTENSÃO

Universidade Federal Rural do Rio de Janeiro

Pró-Reitoria de Pesquisa e Extensão

Curso de Extensão

Curso: Python para Data Science

Prof. Sergio Manuel Serra da Cruz

Prof. Raimundo Macário Costa

Python Básico



PET-SI



Python Básico

Variáveis e tipos de dados

Atribuição de variáveis

Cálculos com variáveis

Tipos e Conversão de tipos **Strings**

Operações e métodos das strings

Listas

Selecionar elementos de uma lista

Operações com listas

Métodos das listas

...

Libraries

Import libraries

```
>>> import numpy
>>> import numpy as np
```

Selective import

```
>>> from math import pi
```

pandas
Data analysis

NumPy
Scientific computing

Machine learning

matplotlib
2D plotting

Install Python

ANACONDA
Leading open data science platform
powered by Python

spyder
Free IDE that is included
with Anaconda

jupyter
Create and share
documents with live code,
visualizations, text, ...



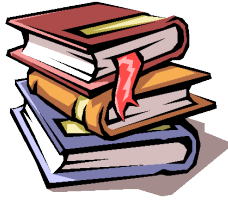
Vamos à prática ...



PET-SI



Bibliografia



PET-SI

1. Wes McKinny, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd edition O'Reilly Media, 2017
2. Mark Lutz, Learning Python, 5th Edition, O'Reilly Media, 2013
3. Jonh Hearty, Advanced Machine Learning with Python. Packt Publishing, 2016.
4. Andreas C. Mueller and Sarah Guido, Machine Learning with Python. O'Reilly Media, 2016.
5. Aurélio Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.
6. John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT, 2015.
7. Manohar Swamynathan, Mastering Machine Learning with Python in Six Steps. Apress, 2017.
8. Michel Bowles, Machine Learning in Python® : Essential Techniques for Predictive Analysis. John Wiley & Sons, Inc. 2015.

